

На правах рукописи



Копосов Александр Сергеевич

**МАТЕМАТИЧЕСКОЕ И АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДЛЯ
ОБРАБОТКИ СЛУЧАЙНЫХ ДАННЫХ С ОГРАНИЧЕННОЙ
ОБЛАСТЬЮ РАССЕЯНИЯ**

Специальность 05.13.01 – Системный анализ, управление и обработка
информации (связь и информатизация)

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Новосибирск – 2016

Работа выполнена на кафедре Радиоэлектроники информационных систем в Федеральном государственном автономном образовательном учреждении высшего профессионального образования Уральском федеральном университете имени первого Президента России Б.Н. Ельцина (ФГАОУ ВПО УрФУ) в городе Екатеринбурге.

Научный руководитель: **Поршнев Сергей Владимирович**
Доктор технических наук, профессор

Официальные оппоненты: **Сызранцев Владимир Николаевич**,
доктор технических наук, профессор,
заведующий кафедрой машин и оборудования
нефтяной и газовой промышленности ФГБОУ
ВО «Тюменский индустриальный
университет»

Кумков Сергей Иванович,
кандидат технических наук, доцент, старший
научный сотрудник отдела динамических
систем Института математики и механики
ФГБУ «Уральское отделение Российской
академии наук»

Ведущая организация: ФГБОУ ВО «Новосибирский государственный
технический университет»

Защита состоится 09 марта 2017 г. в 15:00 часов на заседании диссертационного совета Д 219.005.03 при Федеральном государственном бюджетном образовательном учреждении высшего образования «Сибирский государственный университет телекоммуникаций и информатики» по адресу 630102, г. Новосибирск, ул. Кирова, 86, ауд. 625.

С диссертацией можно ознакомиться в библиотеке Сибирского государственного университета телекоммуникаций и информатики по адресу 630102, г. Новосибирск, ул. Кирова, 86 и на сайте университета, http://www.sibsutis.ru/science/postgraduate/dis_sovets/

Автореферат разослан «____» _____ 201_ г.

Ученый секретарь
диссертационного совета
Д 219.005.03

к.т.н.



Полетайкин Алексей Николаевич

Общая характеристика работы

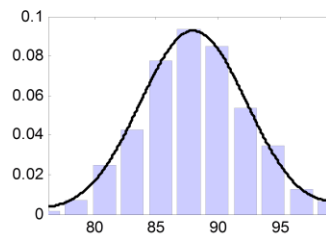
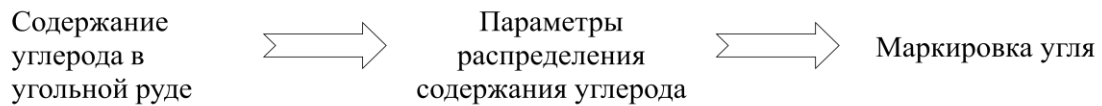
Актуальность темы диссертационного исследования определяется непрерывно возрастающими требованиями к достоверности обработки и интерпретации технологической информации для управления производственными процессами и качеством (проблема Statistical Quality Control).

С математической точки зрения задача обработки технологической информации относится к задаче оценивания функций распределений (ФР) и плотностей вероятностей (ПВ) эмпирических зависимостей по некоторой случайной выборке данных. Для решения этой задачи сегодня разработано большое число различных методов, основанных на подходах параметрической и непараметрической статистики. Однако, только относительно небольшая часть данных методов доведена до законченных программных реализаций, включенных в виде соответствующих инструментов в современные математические пакеты (MATLAB, Statistica, Statgraphics, Mathematica, Maple и др.) и программные библиотеки ряда языков программирования (C++, Python, R и др.)

Анализ систем, используемых в различных отраслях промышленности (угледобывающей, деревообрабатывающей, нефте- и газотранспортной, связи, приборостроительной и др.), показывает, что порождаемая ими технологическая информация с математической точки зрения представляют собой случайные числа, область рассеяния которых является ограниченной (например, время работоспособности приборов, диаметры бревен, концентрации углерода в углях, число пакетов, передаваемых в компьютерных сетях, в течении выбранного временного интервала и др.).

Приведем некоторые примеры подобных задач.

1. Определение параметров распределения содержания углерода в коксующихся углях при торгово-промышленной маркировке угля, которое, исходя из объективной оценки, ограничено снизу примерно 70% и сверху 100%. Данный факт обусловлен тем, что на месторождениях тех или иных количества добывают угли с разной степенью углефикации в указанных пределах. При добыче, отгрузке и транспортировке угля происходит его перемешивание, что может рассматриваться как случайный процесс в ограниченной области рассеяния. Затем происходит многократный забор проб угля. В итоге, получают случайную последовательность, состоящую из значений процентов содержания углерода в пробах угольной руды. На основании параметров данного распределения происходит маркировка угля, т.е. определяется цена угольной руды. Принимая во внимание, что содержание углерода в угле определяет его сортность и, следовательно, цену, понятно, что решение данной задачи имеет важное экономическое значение (рисунок В.1).



Марки угля	Буквенное обозначение марок	содержание углерода С, %
Бурые	Б	76 и менее
Длиннопламенные	Д	76
Газовые	Г	83
Жирные	Ж	86
Коксовые	К	88
Отощенно-спекающиеся	ОС	89
Тощие	Т	90
Антрациты	А	91 и более

Рисунок В.1. К объяснению технологии торгово-промышленной маркировки углей

2. Торгово-промышленная сортировка древесного сырья в деревообрабатывающей промышленности, в частности, сортировка бревен по диаметрам, точность которой определяет эффективность использования древесины при оцилиндровке, и, следовательно, лесопильного производства в целом. Данная сортировка проводится на основе информации о значениях диаметров комлей деревьев данного штабеля (рисунок В.2).

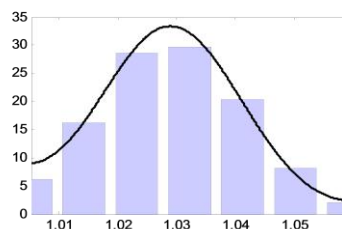
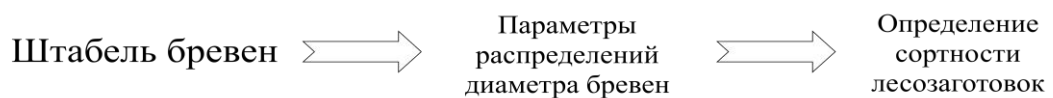


Рисунок В.2. К объяснению технологии торгово-промышленной маркировки лесоматериалов

3. Оценивание прочностной надежности элементов и объектов нефтегазового оборудования с целью определения вероятности безотказной работы нефтепроводов, зависящей от напряжения в трубопроводе и предельного напряжения в трубопроводе, которые также являются случайными величинами с ограниченной областью рассеяния (ограничены слева нулем). При наличии оценки вероятности безотказной работы оказывается возможным вычислить ресурс работы нефтепровода и оценить безопасность технических систем (рисунок В.3).

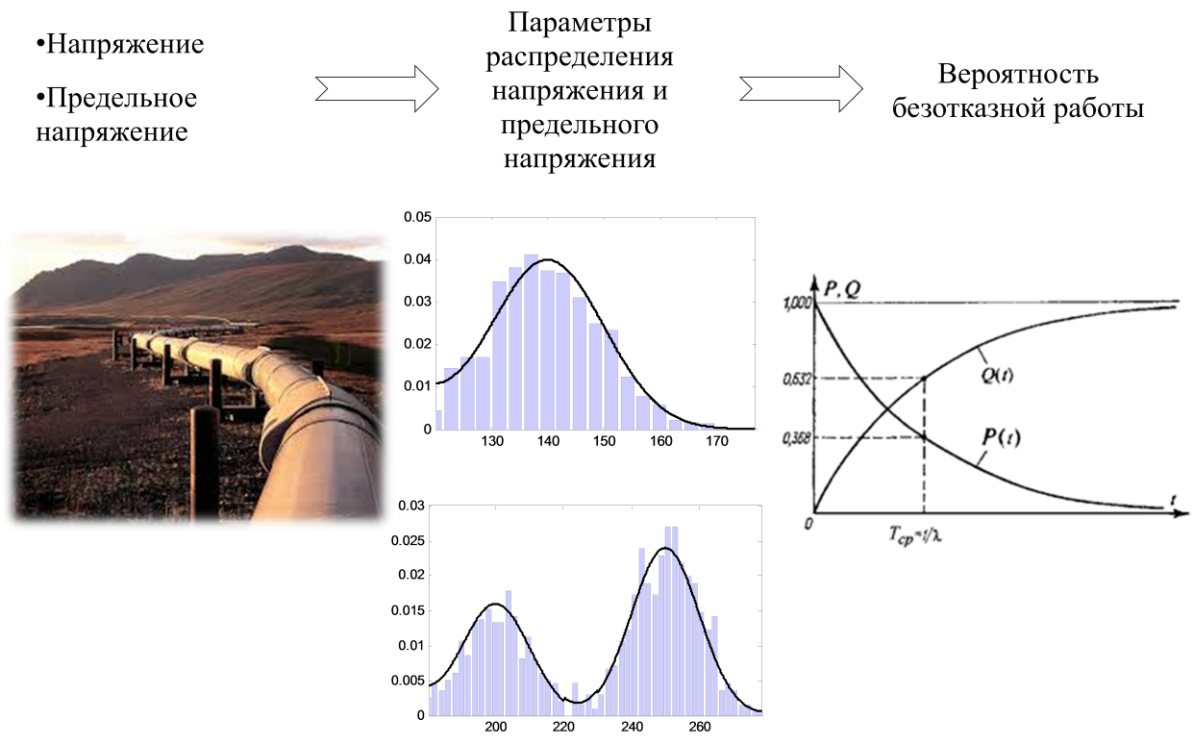


Рисунок В.3. К постановке задачи оценивания вероятности безаварийной работы нефтепровода

4. Оценивание статистических свойств информационных потоков, передаваемых в телекоммуникационных сетях, важность которой обусловлено постоянно увеличением объемов трафика и типов устройств, генерирующих трафик (рисунок В.4).

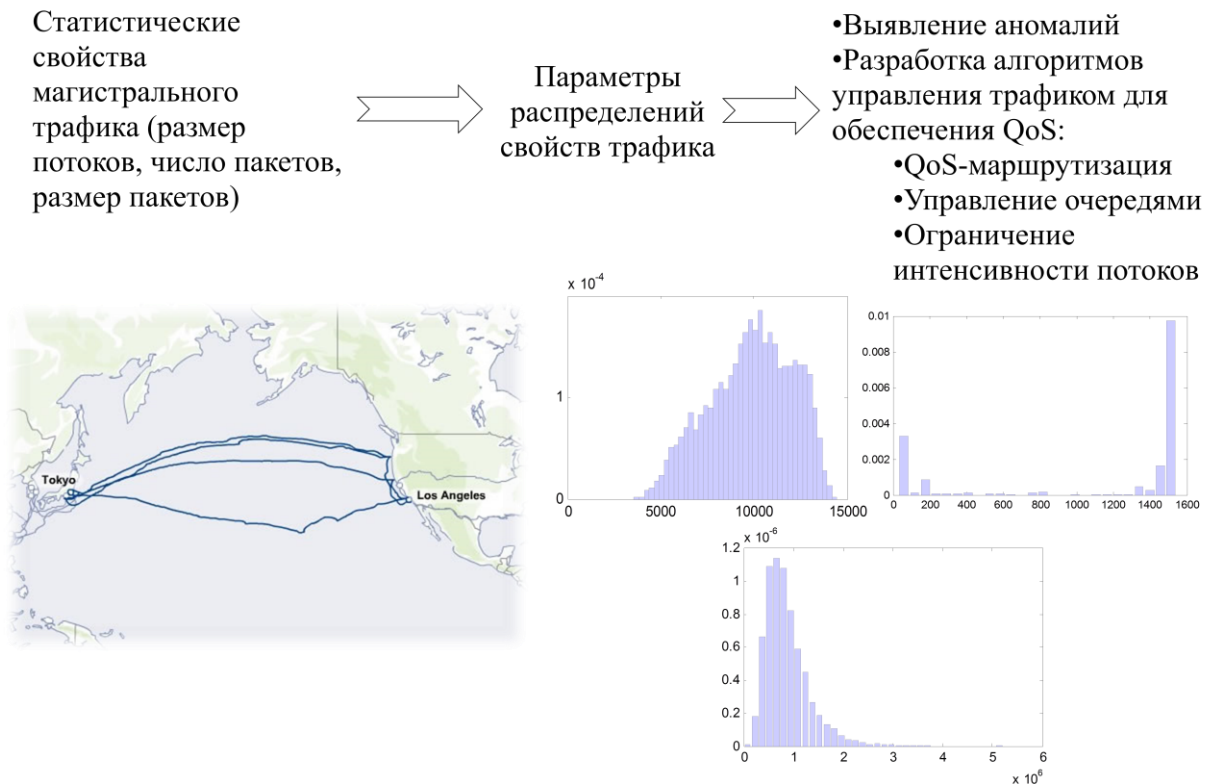


Рисунок В.4. К постановке задачи анализа статистических свойств Интернет-трафика

Здесь для эффективного проектирования телекоммуникационных сетей нового поколения необходимо понимать особенности информационных

потоков и в соответствии с ними использовать адекватные алгоритмы эффективного управления потоками трафика, например: QoS-маршрутизация, управление очередями и планирования обслуживания пакетов, ограничение интенсивности потоков и др.

Теория случайных процессов с ограниченной областью рассеяния (ООР), порождающих соответствующие выборки случайных величин, была построена в работах А. Эйнштейна и М. Смолуховского, в которых они предложили математические модели ФР и ПР случайных величин данного типа. Однако при оценке параметров этих моделей по экспериментальным данным возникает необходимость решения той или иной системы нелинейных уравнений с помощью какого-либо известного итерационного метода (метод Ньютона, симплекс-метод и т.д.). Их сходимость, как известно, определяется удачным выбором начального приближения, сделать который в многомерном случае оказывается не всегда возможным. Данное обстоятельство, с нашей точки зрения, явилось одной из причин недостаточного широкого использования моделей Эйнштейна и Смолуховского в задачах обработки информации.

В то же время сегодня разработаны методы решения экстремальных задач, основанные на случайном поиске (генетические алгоритмы, муравьиные алгоритмы и т.д.), сходимость которых к истинному решению, как утверждается, не зависит от начального приближения. Однако, не существует алгоритмов их использования в задаче аппроксимации эмпирических зависимостей и распределений случайных величин с ООР.

В этой связи оказывается актуальной разработка научно-обоснованных алгоритмов оценивания параметров ФР и ПВ случайных величин с ООР с помощью генетических алгоритмов (ГА), объединяющих параметрические и непараметрические методы оценивания, и их программных реализаций.

Объектом исследования являются методы анализа и обработки технологической информации.

Предмет исследования

Алгоритмическое и математическое обеспечение для оценивания ФР и ПР случайных величин с ООР.

Целью работы является разработка математического и алгоритмического обеспечения для обработки технологической информации, представляющего собой СПООР, на основе использования методов параметрической и непараметрической статистики и эвристических методов.

Для достижения поставленной цели сформулированы и решены следующие **основные задачи исследования**:

1. Анализ особенностей одномодальных распределений СПООР на основе статистического моделирования.

2. Научное обоснование математического и алгоритмического обеспечения для оценивания параметров распределений СПООР, обеспечивающего:

- оценку параметров одномодальных распределений с ООР на основе ГА;

- оценку параметров двумодальных случайных величин с ООР на основе ГА и аппроксимации Розенблатта-Парзена.

3. Разработка программной библиотеки, обеспечивающей практическое использование математического и алгоритмического обеспечения для оценивания параметров ФР и ПВ случайных величин с ООР.

Методы исследования

В работе использованы методы математической статистики, вычислительной математики, системного анализа, имитационного моделирования, ГА.

Теоретико-методологической базой исследований, проведенных в диссертационной работе, послужили труды В.В. Курейчика, Л.А. Гладкова, Х. Гулда, Я. Тобочника, А.Н. Колмогорова, В.Р. Матвиевского, А.И. Орлова, В.А. Симахина, В.Н. Сызранцева, С.Л. Голофаста, А. Эйнштейна, М. Смолуховского, Э. Парзена, М. Розенблатта.

Результаты, выносимые на защиту:

1. Научно-обоснованное математическое и алгоритмическое обеспечение для оценивания ФР и ПВ случайных величин с ООР, включающее в себя:

- методику оценки параметров одномодальных распределений с ООР на основе ГА;
- комплексную методику оценки параметров двумодальных случайных величин с ООР на основе ГА и аппроксимации Розенблатта-Парзена.

Соответствует п. 4 паспорта специальности: Разработка методов и алгоритмов решения задач ... обработки информации; п 5: Разработка специального математического и алгоритмического обеспечения систем ... обработки информации.

2. Алгоритм нахождения оптимального значения параметра размытости h^* аппроксимации Розенблатта-Парзена, адаптированный для дискретных целочисленных случайных последовательностей. *Соответствует п. 4 паспорта специальности: Разработка методов и алгоритмов решения задач ... обработки информации; п 5: Разработка специального математического и алгоритмического обеспечения систем ... обработки информации.*

3. Алгоритм нахождения оптимального значения параметра размытости h^* аппроксимации Розенблатта-Парзена, учитывающий особенности машинного нуля при вычислении на ЭВМ. *Соответствует п. 4 паспорта специальности: Разработка методов и алгоритмов решения задач ... обработки информации; п 5 паспорта специальности: Разработка специального математического и алгоритмического обеспечения систем ... обработки информации.*

4. Программный пакет, реализующий математическое и алгоритмическое обеспечение для оценивания ФР и ПВ случайных величин с ООР. *Соответствует п 5 паспорта специальности: Разработка*

специального математического и алгоритмического обеспечения систем ... обработки информации

Научную новизну исследования представляют следующие результаты исследования:

1. Научно-обоснованное математическое и алгоритмическое обеспечение для оценивания ФР и ПВ случайных величин с ООР, включающее в себя:

- методику оценки параметров одномодальных распределений с ООР на основе ГА;
- комплексную методику оценки параметров двумодальных случайных величин с ООР на основе ГА и аппроксимации Розенблатта-Парзена.

2. Алгоритм нахождения оптимального значения параметра размытости h^* аппроксимации Розенблатта-Парзена, адаптированный для дискретных целочисленных случайных последовательностей.

3. Алгоритм нахождения оптимального значения параметра размытости h^* аппроксимации Розенблатта-Парзена, учитывающий особенности машинного нуля при вычислении на ЭВМ.

Теоретическую значимость представляют следующие результаты работы:

1. Модифицированная для целочисленных последовательностей формула для информационного функционала $\varphi(h)$, используемого в аппроксимации Розенблатта-Парзена, в которой решена проблема немонотонности $\varphi(h)$, что обеспечивает более точную оценку оптимального значения параметра размытости h .

2. Результаты исследований особенностей случайных блужданий в ООР.

3. Формулировки задачи оценивания параметров одномодальных и двумодальных распределений с ООР в терминах ГА.

Практическая значимость исследования

1. Обоснованы рекомендации по выбору областей поиска значений параметров одномодальных распределений с ООР с помощью ГА.

2. Обоснованы рекомендации по выбору областей поиска значений параметров двумодальных распределений с ООР с помощью ГА и аппроксимации Розенблатта-Парзена.

3. Разработана программная библиотека, обеспечивающая практическое использование математического и алгоритмического обеспечения для оценки ФР и ПВ СПООР, на которую получено свидетельство о государственной регистрации программы для ЭВМ № 2016614275.

Полученные в диссертации результаты используются в ООО «Эйконикс», Уральский Банк ПАО «Сбербанк России», ОАО «АйТи-Мед», АО «ЭР-Телеком Холдинг». Основные результаты работы внедрены в учебный процесс Института радиоэлектроники и информационных

технологий Уральского федерального университета им. первого Президента России Б.Н. Ельцина при подготовке бакалавров и магистров по направлению «Информационные системы и технологии»

Апробация работы. Материалы работы докладывались на следующих научных конференциях: Международной научно-практической конференции: «Инновационное развитие: физико-математические и технические науки», Москва, 28 августа 2014 г.; Международной научно-практической конференции «Новое слово в науке: перспективы развития», Чебоксары, 10 сентября 2014 г.; Международной IEEE Сибирская конференция по управлению и связи SIBCON-2015, Омск, 21-23 мая 2015 г.; Международной Крымской конференции «СВЧ-техника и телекоммуникационные технологии» (КрыМиКо'2015), Севастополь, 6-12 сентября 2015; Международной IEEE конференции AICT, Ростов-на-Дону, 14-16 октября 2015 г.; Международная научная конференция «Современные методы прикладной математики, теории управления и компьютерных технологий», Воронеж, 20-26 сентября 2016 г.

Публикации. По теме диссертации опубликовано 19 работ, отражающих основные положения исследования, среди которых 11 статей в журналах, рекомендованных ВАК РФ, из них одна статья вышла в переводной версии журнала, 7 текстов докладов в материалах международных научно-практических конференций. Получено свидетельство о государственной регистрации программы. В работах, опубликованных в соавторстве, лично соискателю принадлежат: исследование особенностей случайных блужданий в ООР; разработка математического и алгоритмического обеспечения для оценивания параметров ФР и ПР с помощью генетических алгоритмов и аппроксимации Розенблатта-Парзена.

Структура и объем работы

Диссертация состоит из введения, четырех глав, заключения, списка литературы из 127 наименований, 8 приложений, содержит 68 рисунков и 39 таблиц. Основной текст работы составляет 130 страниц, общий объем – 185 страниц.

Основное содержание работы

Во **введении** изложена общая характеристика работы, обоснована ее актуальность, сформулирована цель работы, отражена научная новизна и практическая ценность.

В **первой главе** проведен анализ свойств одномодальных и двумодальных распределений с ООР в соответствии с подходом на работах А. Эйнштейна и М. Смолуховского, базирующимся на использовании модели случайного блуждания броуновской частицы в ограниченной области и показано, что использование для аппроксимации ФР данных процессов усеченного нормального закона приводит к возникновению значимых погрешностей.

Проведен анализ постановки задач оценивания параметров одномодальных и двумодальных ФР случайных последовательностей с ООР

результаты которого позволили выдвинуть гипотезы о целесообразности использования в задаче восстановления: 1) одномодальной ФР случайной последовательности ГА; 2) двумодальной ФР случайной последовательности ГА. Сформулированы основные задачи диссертационного исследования.

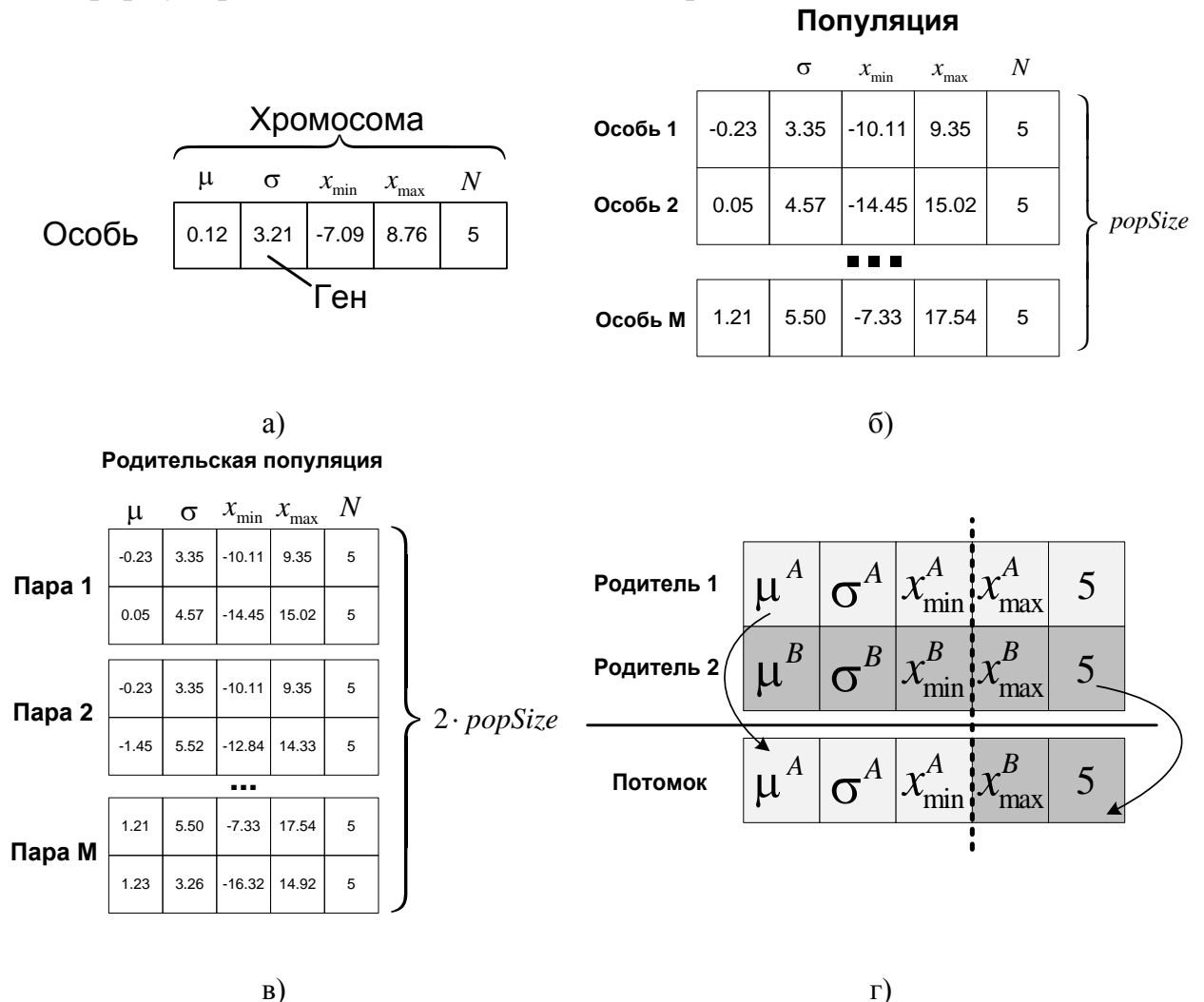


Рис. 1. К установлению соответствия между параметрами распределения и параметрами ГА: а – представление особи, б – пример популяции, в – родительская популяция, полученная в результате селекции, г – пример одноточечного кроссовера

Во **второй главе** обоснована методика использования ГА для оценивания параметров одномодальных ФР случайной последовательности с ООР.

1. Дана постановка задачи оценивания параметров одномодальных ФР случайной последовательности в терминах ГА (рис. 1).

2. Обоснован выбор начальных значений и областей поиска следующих параметров ГА: границ области изменения генов хромосомы $\langle \mu, \sigma, x_{\min}, x_{\max}, N \rangle$; размера популяции; способа формирования начальной популяции; функции приспособленности особи $\Phi(P_i) = \Phi(\{\mu_i, \sigma_i, x_{\min i}, x_{\max i}, N_i\})$; количества особей, подвергаемых процедуре мутации; оператора мутации; оператора селекции; оператора кроссовера.

Блок-схемы вычислительных алгоритмов представлены на рис. 2, 3.

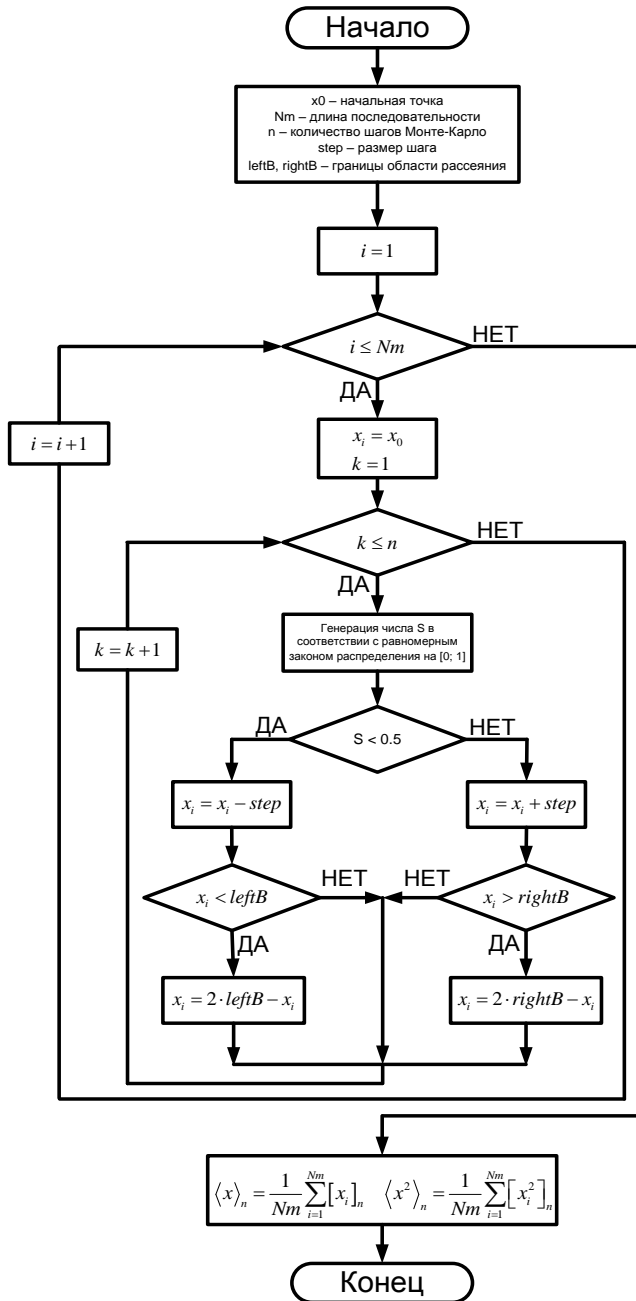


Рис. 2. Блок-схема моделирования случайного блуждания в ограниченной области

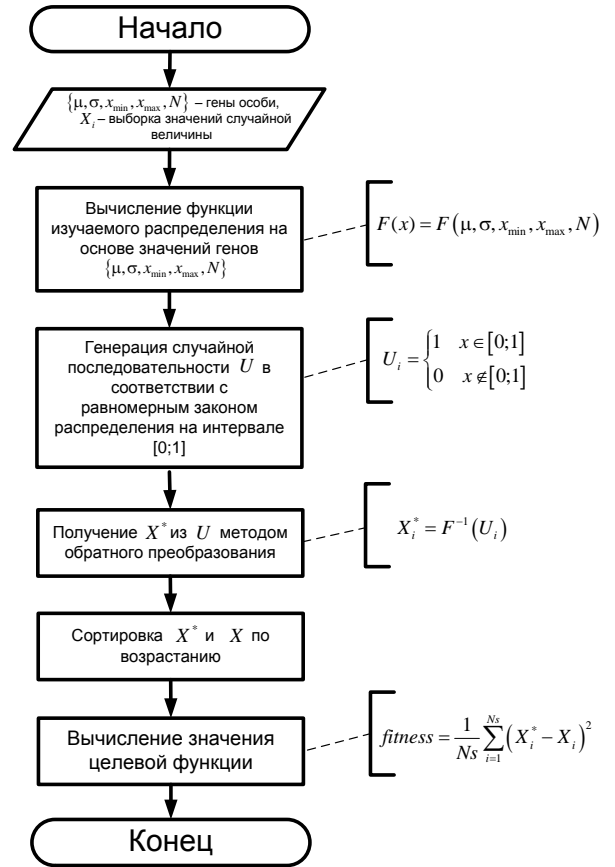


Рис. 3. Блок-схема алгоритма вычисления функции приспособленности: X – анализируемая случайная последовательность

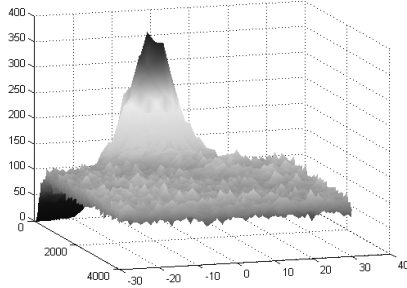
Типичные результаты моделирования случайных блужданий в ООР представлены на рис. 4, из которого видно:

1. Для случая $x_0 = 0$ и $x_2 = -x_1$ плотность распределения при $n \rightarrow \infty$ стремится к равномерному закону распределения с центром в точке $x_0 = 0$ и областью рассеяния $2l = x_2 - x_1$ (рис. 4а, б).

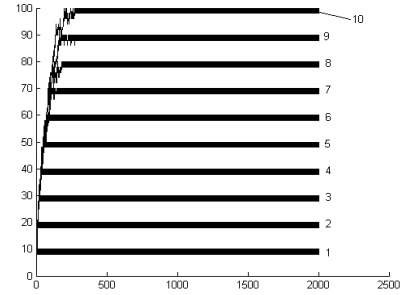
2. Для случая $x_0 = 0$ и $x_2 = -x_1$ зависимость $\langle x^2 \rangle_n$ при $n \rightarrow \infty$ стремится к некоторому предельному значению, определяемому размером области рассеяния $2l$ (рис. 4в, г).

3. Для случая $|x_1| \neq |x_2|$ плотность распределения при $n \rightarrow \infty$ в не зависимости от координаты начальной точки x_0 стремится к равномерному закону распределения с областью рассеяния $2l = x_2 - x_1$ и средним значением $\bar{x} = x_1 + (x_2 - x_1)/2 = (x_1 + x_2)/2$ (рис. 4д).

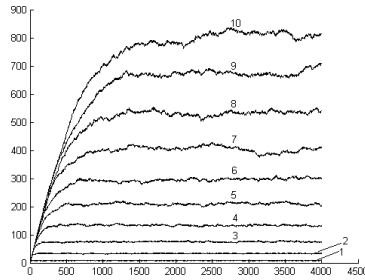
4. Для случая $|x_1| \neq |x_2|$ зависимость $\langle x^2 \rangle_n$ при $n \rightarrow \infty$ вне зависимости от координаты x_0 стремится к некоторому предельному значению, определяемому размером области рассеяния $2l$ (рис. 4е).



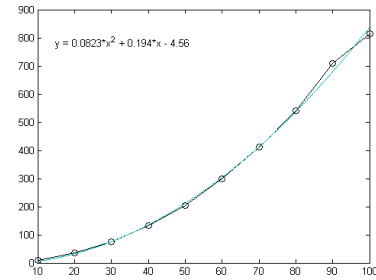
а) Поверхность, задаваемая функцией $Nm \cdot P_n(m)$: $x_0 = 0$, $x_1 = -30$, $x_2 = 30$, $l = 30$, $Nm = 3000$, $n = 50, 60, \dots, 4000$



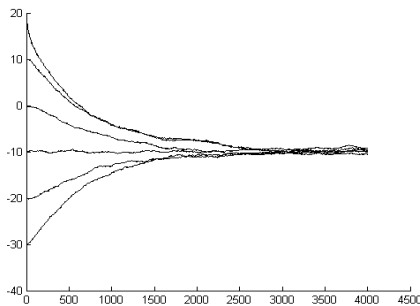
б) Зависимость размаха случайных блужданий от номера шага Монте-Карло: $x_0 = 0$, $x_2 = -x_1$, $|x_1| = |x_2| = l = 10, 20, \dots, 100$, $Nm = 3000$, $n = \overline{1, 2001}$



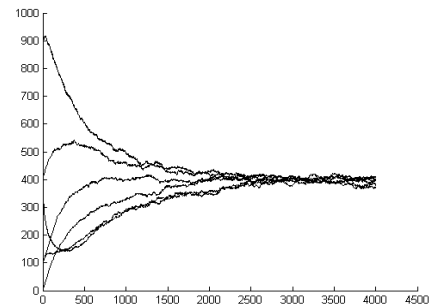
в) Зависимость $\langle x^2 \rangle_n$: $x_0 = 0$, $x_2 = -x_1$, $|x_1| = |x_2| = 2l = 10, 20, \dots, 100$, $Nm = 3000$, $n = \overline{1, 4000}$



г) Зависимость $\langle x^2 \rangle_{4000}$ от $2l$: $x_0 = 0$, $Nm = 3000$



д) Зависимость $\langle x \rangle_n$: $x_0 = -30, -20, \dots, 20$, $x_1 = -40$, $x_2 = 30$, $2l = 70$, $Nm = 3000$, $n = \overline{1, 4000}$



е) Зависимость $\langle x^2 \rangle_n$: $x_0 = -30, -20, \dots, 20$, $x_1 = -40$, $x_2 = 30$, $2l = 70$, $Nm = 3000$, $n = \overline{1, 4000}$

Рис. 4. Результаты статистических исследований: Nm – число броуновских частиц, n – число шагов случайных блужданий, $\langle x \rangle_n$ – координата усредненной по ансамблю реализаций частицы, $\langle x^2 \rangle_n$ – квадрат смещения усредненной по ансамблю реализаций частицы

Результаты статистического моделирования далее были подтверждены соответствующими аналитическими расчетами, представленными в Разделе 2.2.2.

На основе полученных результатов были сформулированы следующие рекомендации по выбору начальных параметров ГА при их использовании в задаче оценивания одномодальной ФР случайной последовательности $x_i, i=1, M$ с ООР:

1. Нулевое приближение параметра x_0 нужно выбирать из отрезка $[\tilde{x}_{\min}, \tilde{x}_{\max}]$, где $\tilde{x}_{\min}, \tilde{x}_{\max}$ – минимальное и максимальное значения случайной последовательности $x_i, i=1, M$.

2. Область поиска истинного значения параметра x_0 следует ограничить отрезком $[\tilde{x}_{\min}, \tilde{x}_{\max}]$.

3. Нулевое приближение параметра $\tilde{\sigma}$ нужно оценивать по экспериментальному значению области рассеяния $2l = \tilde{x}_{\max} - \tilde{x}_{\min}$ и зависимости $\langle x^2 \rangle_n - \langle x^2 \rangle_0 = f(2l)$ (рис. 3в).

4. Область поиска истинного значения параметра σ следует ограничить отрезком $[0, \tilde{\sigma}]$.

Для обоснования выбора функции приспособленности особи $P_i = \{\mu_i, \sigma_i, x_{\min i}, x_{\max i}, N_i\}$ ГА проведен анализ точности оценок параметров одномодальных ФР случайных последовательностей с ограниченной областью рассеяния в соответствии с алгоритмом, блок-схема которого представлена на рис. 2. В ходе проведенных исследований были изучены нормальные распределения с ООР со следующим набором параметров: $N=5$, $\sigma=10$, $x_{\min}=-l$, $x_{\max}=l$, где $2l$ – размер области рассеяния, для $numL=4$ различных значений интервала рассеяния – $2l=2\sigma, 3\sigma, 5\sigma, 10\sigma$ и $numMu=3$ различных положений центра распределения относительно границ интервала: $\mu=0, l/2, l$, обеспечивающий диапазон всех возможных вариантов расположения центра распределения и отношения размера области рассеяния к СКО распределения. Число рассмотренных вариантов сочетаний значений параметров $numSeq$ равнялось 12.

В качестве эталона для функции приспособленности $P_i = \{\mu_i, \sigma_i, x_{\min i}, x_{\max i}, N_i\}$ использовалась экспериментальная выборка X значений изучаемой случайной последовательности, сгенерированной в соответствии с алгоритмом, блок-схема которого представлена на рис. 2, и следующие настройки ГА: селекция – $S = \langle \text{равномерная; турнирная; на основе рулетки} \rangle$; мутация – $M = \langle \text{адаптивная} \rangle$ (выбрана, т.к. в задаче присутствуют ограничения); кроссовер – $C = \langle \text{одноточечный; двухточечный; усредненный; разбросанный, эвристический} \rangle$; доля кроссовера – $Cfr = \langle 0,3; 0,6; 0,9 \rangle$; доля мутации $1 - Cfr$; размер популяции: $P = \langle 5, 10, 25 \rangle$. Общее число использованных в экспериментах комбинаций настроек $numT$

составило 135. Для каждого набора настроек было выполнено $numExp = 10$ независимых статистических испытаний, в ходе которых генерировались случайные последовательности $\{x_i\}^{(k)}, i = \overline{1, Ns}$ ($Ns = 500$) в соответствие с заданным законом распределения и, далее, следуя алгоритму, представленному на рис. 3, оценивались параметры распределения. Кроме того по каждому k -ому независимому испытанию оценивалась величина, характеризующее отличие между теоретической и экспериментальной функциями распределения:

$$[\Delta_{integr}]^{(k)} = \frac{\sum_i^{Ns_k} \left(F_{teor}(x_i, \mu, \sigma, x_{\min}, x_{\max}, N) - [F_{pract}(x_i, \tilde{\mu}, \tilde{\sigma}, \tilde{x}_{\min}, \tilde{x}_{\max}, N)]^{(k)} \right)^2}{\sum_i^{Ns_k} F_{teor}(x_i, \mu, \sigma, x_{\min}, x_{\max}, N)^2}.$$

где

N – число мнимых источников (равное 5);

$\tilde{\mu}, \tilde{\sigma}, \tilde{x}_{\min}, \tilde{x}_{\max}$ – экспериментально вычисленные параметры распределения;

$\mu, \sigma, x_{\min}, x_{\max}$ – теоретические параметры распределения;

k – номер испытания;

Ns_k – число элементов в последовательности.

Общее число запусков ГА составило

$$numTotal = numT \times numSeq \times numExp = 135 \times 12 \times 10 = 16200.$$

Анализ полученных результатов показал, что наилучшие результаты ($\Delta_{integr} = 1 \cdot 10^{-5}; 4 \cdot 10^{-5}; 9 \cdot 10^{-5}$) получены при следующих настройках ГА: селекция: **равномерная** или **турнирная**; мутация: **адаптивная**; кроссовер – **одноточечный** или **эвристический**; доля кроссовера находится в интервале от **0.2** до **0.7**; размер популяции: **25–50**.

На основании полученных результатов предложен следующий алгоритм оценивания параметров случайной последовательности $X_i, i = \overline{1, Ns}$ с ООР:

1. Вычисление оценки координаты левой границы области рассеяния распределения $\tilde{x}_{\min} = \min_i (X_i)$.

2. Вычисление оценки координаты правой границы области рассеяния распределения $\tilde{x}_{\max} = \max_i (X_i)$.

3. Задание количества пар фиктивных источников $N = 5$.

4. Задание интервала ограничения для параметра μ : $[\tilde{x}_{\min}, \tilde{x}_{\max}]$.

5. Задание интервала ограничения для параметра σ : $[0, \sigma_{\max}]$, где σ_{\max} вычисляется по зависимости (рис. 4в).

6. Задание настроек ГА: селекция – равномерная или турнирная; мутация – адаптивная; кроссовер – одноточечный или эвристический; доля кроссовера – в интервале от 0.2 до 0.7; размер популяции – 25–50.

7. Вычисление, в соответствии с ГА, значений параметров μ, σ .

В третьей главе обоснована методика оценивания параметров двумодальных ПР и ФР случайных последовательностей с ООР. Напомним, что ФР данной случайной величины вычисляются по формуле:

$$F_{2\text{mod}}(x) = F_{LAD}(x, \mu_1, \sigma_1, a_1, b_1) \cdot \alpha + F_{LAD}(x, \mu_2, \sigma_2, a_2, b_2) \cdot (1 - \alpha),$$

где

μ_1 – математическое ожидание первой составляющей;

σ_1 – СКО первой составляющей;

a_1, b_1 – границы области рассеяния первой составляющей;

μ_2 – математическое ожидание второй составляющей;

σ_2 – СКО второй составляющей;

a_2, b_2 – границы области рассеяния второй составляющей;

α – доля первой составляющей в общем распределении.

Для этого сформулирована постановка задачи оценивания параметров двумодальных ПР и ФР случайных последовательностей с помощью ГА (рис. 5).

На первом этапе были использованы классические настройки ГА и рекомендации по выбору начальных значений и границ областей поиска параметров, аналогичные описанным в главе 2.

Было изучено $numSeq=5$ различных двумодальных распределений случайных последовательностей с ООР. Наилучшие результаты работы ГА для каждого набора параметров распределения представлены в табл. 1.

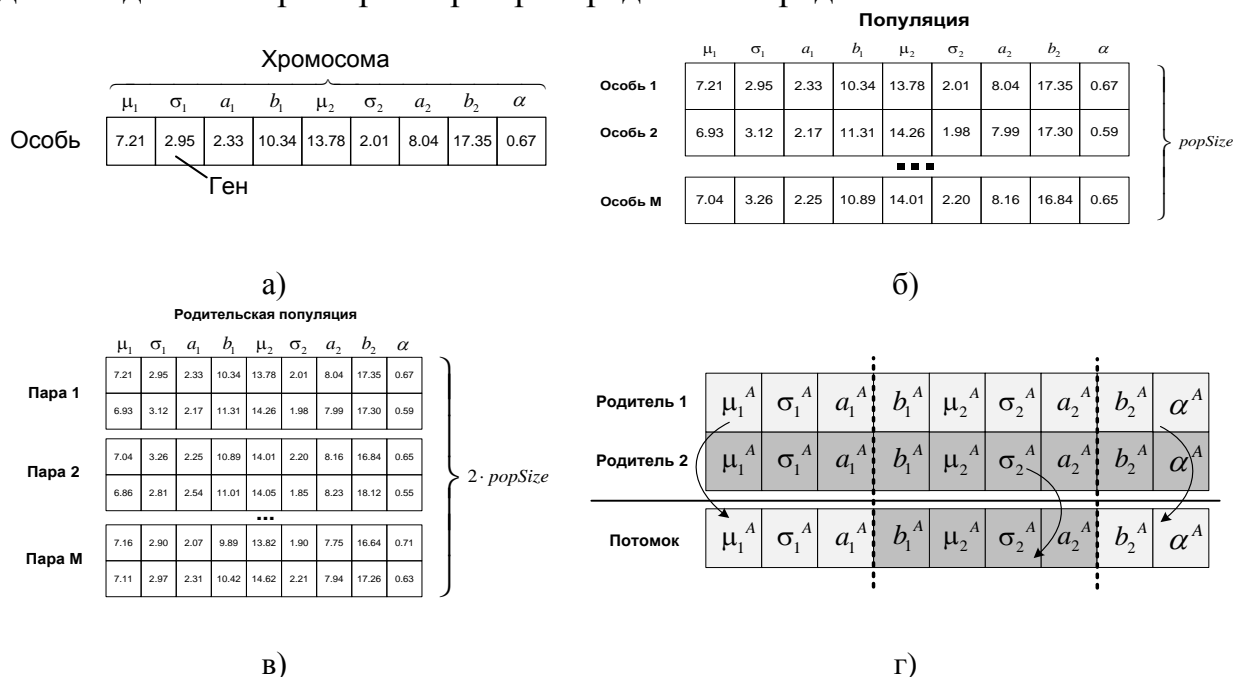


Рис. 5. К установлению соответствия между параметрами распределения и параметрами ГА: а – представление особи, б – пример популяции, в – родительская популяция, полученная в результате селекции, г – пример двухточечного кроссовера

Таблица 1. Точность оценки параметров двумодального распределения с ООР

№ набора параметров	Δ_{integr}
1	$1.118 \cdot 10^{-1}$
2	$9.403 \cdot 10^{-2}$
3	$1.992 \cdot 10^{-2}$
4	$6.801 \cdot 10^{-3}$
5	$1.112 \cdot 10^{-1}$

Из табл. 1 видно, что точность аппроксимации двумодальной ПР случайной величины с ООР для наборов параметров 1, 2 и 5 недостаточно высока, а интегральная погрешность на порядок больше, чем интегральная погрешность, полученная при анализе точности оценок параметров одномодальных распределений. Отмеченное обстоятельство определило необходимость совершенствования данной методики на основе использования аппроксимации Розенблатта-Парзена для получения независимых оценок одного или нескольких параметров двумодального распределения.

Напомним, что для вычисления аппроксимации плотности распределения случайной последовательности в методе Розенблатта-Парзена необходимо определить оптимальное значение параметра размытости h^* , при котором достигается максимум информационного функционала

$$\varphi(h) = \frac{1}{Ns} \sum_{i=1}^{Ns} \ln \left[\frac{1}{(Ns-1) \cdot h} \sum_{j \neq i}^{Ns-1} k_m \left(\frac{x_i - x_j}{h} \right) \right].$$

Найти решение уравнения $d\varphi(h)/dh = 0$ оказывается возможным только численно. При выполнении численных расчетов было обнаружено, что в ряде случаев при малых значений h зависимость $\varphi(h)$ имеет разрывы первого рода, что затрудняет автоматизированный поиск значения h^* . Причина выявленной особенности, как показал дальнейший анализ функции $\varphi(h)$, состоит в том, что при численных расчетах вычисленные значения

некоторых ядерных функций $k \left(\frac{x_i - x_j}{h} \right)$ оказывались меньше машинного

нуля. Свободной от указанного недостатка оказалась предложенная в работе модифицированная формула для информационного функционала

$$\varphi(h) = -\ln(Ns-1) - \ln h + \sum_{i=1}^{Ns} \ln \left[\sum_{j \neq i}^{Ns-1} k_m \left(\frac{x_i - x_j}{h} \right) \right].$$

Также обосновано, что поиск оптимального значения h^* для нормального ядра, ядра Лапласа, ядра Коши, логистического ядра следует, соответственно, искать на интервалах:

$$\left[\left(\frac{\min((x_i - x_j)^2 / 2)}{-\lg m_0 \cdot \ln 10} \right)^{1/2}, \max((x_i - x_j)^2 / 2)^{1/2} \right]$$

$$\left[\frac{\min(|x_i - x_j|)}{-\lg m_0 - \lg 2}, \max(|x_i - x_j|) \right], \left[\frac{\min(x_i - x_j)^2}{1/(\pi \cdot m_0) - 1}, \max(|x_i - x_j|) \right],$$

$$\left[\arg_h \left(2 \cdot \lg \left(1 + e^{-\min(|x_i - x_j|)/h} \right) + \frac{\min(|x_i - x_j|)}{h} \cdot \lg e = -\lg m_0 \right), \max(|x_i - x_j|) \right],$$

а оптимальные значения параметра размытости для ядер с ограниченной областью определения следует вычислять по следующим формулам: ядро Фишера – $h = \max(x_i - x_j) / 2\pi$, ядро Епанечникова – $h = \max(x_i - x_j) / \sqrt{5}$, равномерное, треугольное и квадратичное ядра – $h = \max(x_i - x_j)$.

Предложена методика оценивания параметров случайной величины со смешанным двумодальным распределением на основе итерационного использования аппроксимации Розенблатта-Парзена, реализующаяся следующей последовательностью действий:

1. Вычисление в соответствии с методом Розенблатта-Парзена значений функции $f_{RP}(x)$, аппроксимирующей двумодальную ПР. Наличие значений аппроксимирующей функции позволяет оценить примерный вид исходной функции распределения, а также получить оценки координаты левой границы области рассеяния распределения, описываемой функцией $F_1(x, \mu_1, \sigma_1, a_1, b_1)$:

$$\tilde{a}_1 = \min_i (X_i),$$

и правой границы области рассеяния распределения, описываемой функцией $F_2(x, \mu_2, \sigma_2, a_2, b_2)$:

$$\tilde{b}_2 = \max_i (X_i),$$

а также вычислить оценки значений параметров $\tilde{\mu}_1, \tilde{\mu}_2$ – абсцисс локальных максимумов функции $F_{RP}(x)$.

2. Вычисление с помощью ГА значений параметров $\sigma_1, b_1, \sigma_2, a_2, \alpha$. (Здесь μ_1, μ_2 – оценки, полученные на этапе 1 по аппроксимации Розенблатта-Парзена, которые считаются постоянными.)

3. Вычисление с помощью ГА значений параметров $\mu_1, b_1, \mu_2, a_2, \alpha$. (Здесь σ_1, σ_2 – оценки, полученные на этапе 2, которые считаются постоянными.)

4. Вычисление с помощью ГА значений параметров $\sigma_1, \sigma_2, \alpha$. (Здесь μ_1, b_1, μ_2, a_2 – оценки, полученные на этапе 3, которые считаются

постоянными, в качестве начального приближения параметров $\sigma_1, \sigma_2, \alpha$ используются значения, полученные на этапе 3).

5. Вычисление, используя ГА, значения параметров μ_1, μ_2 . (Здесь $\sigma_1, b_1, \sigma_2, a_2, \alpha$ – оценки, полученные на этапе 4, которые считаются постоянными, в качестве начального приближения параметров μ_1, μ_2 используются значения, полученные на этапе 4)

6. Вычисление, используя ГА, значения параметров $\sigma_1, \sigma_2, \alpha$. (Здесь μ_1, b_1, μ_2, a_2 – оценки, полученные на этапе 5, которые считаются постоянными, в качестве начального приближения параметров $\sigma_1, \sigma_2, \alpha$ используются значения, полученные на этапе 5)

В четвертой главе приведено описание программной библиотеки «ES&RP», обеспечивающей практическое использование разработанного в Главе 2 и Главе 3 математического и алгоритмического обеспечения для оценивания параметров ФР и ПВ СПООР. В состав библиотеки входят 42 функции, в том числе, позволяющие реализовать итерационный алгоритм оценивания параметров функций распределения и плотностей распределения двумодальных случайных величин с ограниченной областью рассеяния, основанный на совместном использовании метода мнимых источников и аппроксимации Розенблатта-Парзена. Для вычисления аппроксимации Розенблатта-Парзена предусмотрена возможность использования следующих ядерных функций: функция Коши, функция Епанчикова, функция Фишера, функция Лапласа, логистическая функция, нормальная функция, квадратичная функция, треугольная функция, равномерная функция.

Основные функции библиотеки:

function [Mu, Sigma, a1, a2, y] = solveBoundDistrGA(x, cx)

Функция, возвращающая

Mu, Sigma, a1, a2 – параметры распределения

y – значения функции плотности распределения для значений аргумента cx

Входные параметры

x – случайная последовательность

cx – значения по оси ОХ для функции плотности

function [D11, D12, alpha1, fPract1,...

D21, D22, alpha2, fPract2,...

D31, D32, alpha3, fPract3,...

D41, D42, alpha4, fPract4,...

D51, D52, alpha5, fPract5, fPR] = solveDoublePeak(arr, cx)

Функция, возвращающая

D#1 – структура, содержащая параметры первого распределения на этапе #

D#2 – структура, содержащая параметры второго распределения на этапе #

alpha# – коэффициент первого распределения на этапе #

fpract# – вектор значений аппроксимации распределения на этапе #

fPR – значения аппроксимации Розенблатта-Парзена

Входные параметры

cx – значения по оси ОХ для плотности распределения

arr – случайная последовательность

function [y, kernel, h, J] = solvePR(arr, cx)

Функция, возвращающая

y – значение аппроксимации Розенблатта-Парзена

kernel – наименование наилучшей ядерной функции

h – оптимальное значение параметра размытости

J – оптимальное значение информационного функционала

Входные параметры

arr – случайная последовательность

cx – значения аргумента для вычисления плотности распределения

Приведены примеры, подтверждающие работоспособность созданной программной библиотеки, при оценивании параметров одномодальных и двумодальных случайных последовательностей с ООР

В **заключении** изложены основные результаты диссертационного исследования.

Приложения содержат таблицы, содержащие результаты исследования оценки параметров распределений.

Основные результаты и выводы

1. Проведен анализ особенностей усеченного нормального распределения и нормального распределения с ООР, позволивший выявить отличия между усеченным нормальным распределением и нормальным распределением с ограниченной областью рассеяния.

2. Проведено исследование особенностей одномодальных распределений случайных последовательностей с ограниченной областью рассеяния на основе статистического моделирования, результаты которого позволили обосновать рекомендации по выбору начальных значений и областей поиска параметров ГА.

3. Разработаны модификации алгоритмов нахождения оптимального значения параметра размытости h^* при использовании аппроксимации Розенблатта-Парзена, в котором учитывается особенность машинного нуля ЭВМ; нахождения оптимального значения параметра размытости h^* при

аппроксимации Розенблатта-Парзена, адаптированный для дискретных целочисленных случайных последовательностей.

4. Разработано математическое и алгоритмическое обеспечение для оценивания функций распределений случайных величин с ООР, включающее в себя:

- методику оценивания параметров одномодальных распределений с ООР на основе ГА;
- комплексную методику оценивания параметров двумодальных распределений случайных величин с ООР на основе итерационного использования ГА и аппроксимации Розенблатта-Парзена.

5. Разработана программная библиотека, обеспечивающая практическое использование математического и алгоритмического обеспечения для оценивания функций распределений случайных величин с ООР.

Основные результаты диссертационного исследования опубликованы в следующих работах

Публикации в изданиях, рекомендованных ВАК РФ

1. Копосов А.С. Опыт использования суперкомпьютера для обработки дампов сетевого трафика магистрального интернет-канала / С.В. Поршневу, Д.А. Божалкин, А.С. Копосов // Информационные технологии. 2016 №1(22). С 42-47.
2. Копосов А.С.. Исследование особенностей потоков сетевого трафика в магистральном интернет-канале / С. В. Поршневу, Д. А. Божалкин, А. С. Копосов // Электросвязь. 2016 №2. С 27-33.
3. Копосов А.С. О выборе математических моделей распределений ограниченных случайных последовательностей [Электр.] / С.В. Поршневу, А.С. Копосов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №10(84). – Режим доступа: <http://ej.kubagro.ru/2012/10/pdf/53.pdf>.
4. Копосов А.С. Аналитическое исследование особенностей случайных блужданий броуновской частицы в ограниченной области рассеяния / С.В. Поршневу, А.С. Копосов // Фундаментальные исследования. – 2013. – № 4 (часть 1). – стр. 57-64.
5. Копосов А.С. Исследование особенностей случайных блужданий броуновской частицы в ограниченной области рассеяния на основе статистического моделирования / С.В. Поршневу, А.С. Копосов // Фундаментальные исследования. 2013. №6 (часть 2). С. 284-290.
6. Копосов А.С. Об особенностях численных оценок параметра размытости плотностей распределений случайных последовательностей в аппроксимации Розенблатта-Парзена / С.В. Поршневу, А.С. Копосов // В мире научных открытий. – 2013 №10(46).
7. Копосов А.С. Использование аппроксимации Розенблатта-Парзена для восстановления функции распределения дискретной случайной величины / С.В. Поршневу, А.С. Копосов // В мире научных открытий. 2013 №10(46).
8. Копосов А.С. Использование аппроксимации Розенблатта-Парзена для восстановления функции распределения непрерывной случайной величины с ограниченным одномодальным законом распределения / С.В. Поршневу, А.С. Копосов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. Краснодар: КубГАУ, 2013. №08(092). IDA [article ID]: 0921308076.
9. Копосов А.С. Методика оценивания параметров случайной величины со смешанным двумодальным законом распределения на основе итерационного использования аппроксимации Розенблатта–Парзена и метода мнимых источников / С.В. Поршневу, А.С. Копосов // Фундаментальные исследования. – 2014. – № 5 (часть 5). – стр. 974-983.
10. Копосов А.С. Методика оценивания параметров случайной величины со смешанным двумодальным законом распределения на основе совместного использования аппроксимации Розенблатта–Парзена, метода мнимых источников и генетических алгоритмов / С.В. Поршневу, А.С. Копосов // Фундаментальные исследования. – 2014. – № 8 (часть 3). – стр. 583-589.
11. Копосов А.С. Методика использования генетических алгоритмов в задаче оценки параметров распределений с ограниченной областью рассеяния / С.В. Поршневу, А.С. Копосов // Современные проблемы науки и образования. 2014. №4. С. 168.
12. Koposov A.S. Quantitative estimation peculiarities of densities blur parameter of random sequences distribution in Parzen-Rosenblatt approximation / S.V. Porshnev, A.S. Kposov // In the World of Scientific Discoveries, Series B, vol. 2, №1, 2014, pp. 120–129.11.

Свидетельства на программу для ЭВМ

13. Поршнев С.В., Копосов А.С. Программная библиотека ES&RP. - Свидетельство о государственной регистрации программы для ЭВМ//Свидетельство о государственной регистрации программы для ЭВМ № 2016614275 (Заявка № 2016611747 Дата поступления 2 марта 2016 г. Дата государственной регистрации в Реестре программ для ЭВМ 20 апреля 2016 г.).

Публикации в других изданиях

14. Копосов А.С. Исследование особенностей потоков сетевого трафика в магистральном интернет-канале: распределения размеров пакетов / С.В. Поршнев, А.С. Копосов, Д.А. Божалкин // Международная научная конференция «Современные методы прикладной математики, теории управления и компьютерных технологий», Воронеж, 20-26 сентября 2016 г

15. Koposov A.S. The research of a network traffic in a backbone internet channel in Control and Communications / S.V. Porshnev, A.S. Koposov, D.A. Bozhalkin // 2015 International Siberian Conference on Control and Communications (SIBCON). PROCEEDINGS RUSSIA, OMSK, MAY 21–23, 2015. doi: 10.1109/SIBCON.2015.7147005: IEEE Catalog Number: 15287637. URL: <http://ieeetpu.ru/musor/sbornik/papers/050fu.pdf> (Дата доступа: 09.06.2015 г.)

16. Koposov A.S. Features of information flows in the backbone Internet-channel: the analysis of the statistical characteristics of the relationship between the number of packets and the time / S.V. Porshnev, A.S. Koposov, D.A. Bozhalkin // 2015 9th International Conference on Application of Information and Communication Technologies, Russia, Rostov-on-Don, 14-16 October 2015.

17. Копосов А.С. Особенности информационных потоков в магистральном интернет-канале: анализ статистических характеристик зависимостей объема пакетов от времени / С.В. Поршнев, А.С. Копосов, Д.А. Божалкин // Международная Крымская конференция «СВЧ-техника и телекоммуникационные технологии» (КрыМиКо'2015), Севастополь, 6-12 сентября 2015.

18. Копосов А.С. Использование генетических алгоритмов в задаче оценки параметров неусеченного нормального распределения на основе усеченной выборки (доклад) / С.В. Поршнев, А.С. Копосов // Теория, техника и экономика сетей связи: сборник научно-технических и методических трудов. Выпуск 10. Екатеринбург: СибГУТИ, 2012. С 390 – 392.

19. Копосов А.С. Методика оценивания параметров случайной величины со смешанным двумодальным законом распределения на основе совместного использования аппроксимации Розенблатта-Парзена, метода мнимых источников и генетических алгоритмов / С.В. Поршнев, А.С. Копосов // Новое слово в науке: перспективы развития: сборник материалов международной научно–практической конференции. Чебоксары, 2014. С. 231-236.

20. Копосов А.С. Методика использования генетических алгоритмов в задаче оценки параметров распределений с ограниченной областью рассеяния / С.В. Поршнев, А.С. Копосов // Международная научно-практической конференция: «Инновационное развитие: физико-математические и технические науки», Москва, 28 августа 2014 г.